

SEQUENCE LISTING

<110> Flint, Dennis

Meyer, Knut

Viitanen, Paul

<120> Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates From Benzoic Acid Glucosides

<130> BC1034 US NA

<140> US 60/216,615

<141> 2000-07-07

<150> 60/216,615

<151> July 7, 2000

<160> 19

<170> Microsoft Office 97

<210> 1

<211> 433

<212> PRT

<213> Arabidopsis thaliana

<400> 1

Met Ser Leu Lys Ile Lys Phe Leu Leu Leu Val Leu Tyr His His
1 5 10 15

Val Asp Ser Ala Ser Ile Val Lys Phe Leu Pro Gly Phe Glu Gly Pro
20 25 30

Leu Pro Phe Glu Leu Glu Thr Gly Tyr Ile Gly Ile Gly Glu Asp Glu
35 40 45

Asn Val Gln Phe Phe Tyr Tyr Phe Ile Lys Ser Glu Asn Asn Pro Lys
50 55 60

Glu Asp Pro Leu Leu Ile Trp Leu Asn Gly Gly Pro Gly Cys Ser Cys
 65 70 75 80

Leu Gly Gly Ile Ile Phe Glu Asn Gly Pro Val Gly Leu Lys Phe Glu
 85 90 95

Val Phe Asn Gly Ser Ala Pro Ser Leu Phe Ser Thr Thr Tyr Ser Trp
 100 105 110

Thr Lys Met Ala Asn Ile Ile Phe Leu Asp Gln Pro Val Gly Ser Gly
 115 120 125

Phe Ser Tyr Ser Lys Thr Pro Ile Asp Lys Thr Gly Asp Ile Ser Glu
 130 135 140

Val Lys Arg Thr His Glu Phe Leu Gln Lys Trp Leu Ser Arg His Pro
 145 150 155 160

Gln Tyr Phe Ser Asn Pro Leu Tyr Val Val Gly Asp Ser Tyr Ser Gly
 165 170 175

Met Ile Val Pro Ala Leu Val Gln Glu Ile Ser Gln Gly Asn Tyr Ile
 180 185 190

Cys Cys Glu Pro Pro Ile Asn Leu Gln Gly Tyr Met Leu Gly Asn Pro
 195 200 205

Val Thr Tyr Met Asp Phe Glu Gln Asn Phe Arg Ile Pro Tyr Ala Tyr
 210 215 220

Gly Met Gly Leu Ile Ser Asp Glu Ile Tyr Glu Pro Met Lys Arg Ile
 225 230 235 240

Cys Asn Gly Asn Tyr Tyr Asn Val Asp Pro Ser Asn Thr Gln Cys Leu
 245 250 255

Lys Leu Thr Glu Glu Tyr His Lys Cys Thr Ala Lys Ile Asn Ile His
 260 265 270

His Ile Leu Thr Pro Asp Cys Asp Val Thr Asn Val Thr Ser Pro Asp
 275 280 285

Cys Tyr Tyr Tyr Pro Tyr His Leu Ile Glu Cys Trp Ala Asn Asp Glu
 290 295 300

Ser Val Arg Glu Ala Leu His Ile Glu Lys Gly Ser Lys Gly Lys Trp
 305 310 315 320

Ala Arg Cys Asn Arg Thr Ile Pro Tyr Asn His Asp Ile Val Ser Ser
 325 330 335

Ile Pro Tyr His Met Asn Asn Ser Ile Ser Gly Tyr Arg Ser Leu Ile
 340 345 350

Tyr Ser Gly Asp His Asp Ile Ala Val Pro Phe Leu Ala Thr Gln Ala
 355 360 365

Trp Ile Arg Ser Leu Asn Tyr Ser Pro Ile His Asn Trp Arg Pro Trp
 370 375 380

Met Ile Asn Asn Gln Ile Ala Gly Tyr Thr Arg Ala Tyr Ser Asn Lys
 385 390 395 400

Met Thr Phe Ala Thr Ile Lys Gly Gly His Thr Ala Glu Tyr Arg
405 410 415

Pro Asn Glu Thr Phe Ile Met Phe Gln Arg Trp Ile Ser Gly Gln Pro
420 425 430

Leu

<210> 2

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<223> primer

<400> 2

tcatgacctc tatcgtaag tttcttcc

28

<210> 3

<211> 6

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<223> primer

<400> 3

tcatga

6

<210> 4

<211> 25

<212> DNA

<213> Artificial Sequence

```

<220>
<221> misc_feature
<223> primer

<400> 4
gtcgacttac aggggttggc cactg                                25

<210> 5
<211> 1302
<212> DNA
<213> Arabidopsis thaliana

<400> 5
atgagtttga aaataaaagtt tctgcttctg ctgttcttgt atcatcatgt tgattctgcc   60
tctatcgta agtttcttcc tggtttgaa ggccctcttc ctttcgaaact tgaaaccggg   120
tacattggta ttggtgagga cgagaatgtg caattttct actatttcat caaatctgaa   180
aacaatccaa aagaagatcc tcttcttata tggtaaatg gaggacctgg atgttcttgt   240
cttggtggta ttattttga gaacggaccg gtggggttga agtttgaggt gttcaacggg   300
agtgcctt cttgttctc tactacatat tcatggacaa agatggcaaa cattatattc   360
ttggatcagc cagtaggatc tggcttctcc tactcaaaaa ctccaattga taaaactggt   420
gacataagtg aagtaaagag gacccatgag tttcttcaaa agtggctaag caggcatcca   480
caatatttct ccaacccttt atatgttagtt ggagattctt attccggtat gattgtcccc   540
gccctcggttca aagaaatctc acaaggaaat tatatatgtt gcgaacctcc tataaatcta   600
cagggttata tgcttgaaa ccctgtaca tatatggact ttgaacaaaa cttccgcatt   660
ccatatgctt atggatggg attaatctcc gacgaaatct atgagccaaat gaagagaatc   720
tgcaacggaa attattacaa tgtggatcca tctaacacac aatgtttgaa acttactgaa   780
gaatatcata agtgcactgc caaaataat atccatcaca tattaacacc agattgcgt   840
gtaaccaatg taacatctcc tgattgttat tattatccat atcatctcat tgaatgttgg   900
gctaacgacg agagcgatcg cgaagcttt catattgaaa aggtagttaa aggaaaatgg   960
gcgcgatgta atcggactat tccatacaat cacgacattg taagcagcat accatatcac   1020
atgaataaca gcatcagtgg ataccgatct cttatataca gtggatgatca cgacatcgcg   1080
gtcccttttc ttgcaactca agcctggata agatctctca attactcccc cattcataac   1140
tggaggccat ggatgataaa caatcaaatac gctggataca cgagagctt ttccaataag   1200
atgacatttgc tcaactatcaa aggaggatggcacacggcag agtatacgacc aaacgagacc   1260

```

tttatcatgt tccaaaggtg gatcagtggc caaccctgt aa	1302
<210> 6	
<211> 1256	
<212> DNA	
<213> Arabidopsis thaliana	
<400> 6	
tcatgacctc tatcgtaag tttcttcctg gtttgaagg ccctttcct ttcgaacttg	60
aaaccgggta cattggatt ggtgaggacg agaatgtgca attttctac tatttcatca	120
aatctgaaaa caatccaaaa gaagatcctc ttcttatatg gttaaatgga ggacctggat	180
gttcttgtct tgggtgatt attttgaga acggaccggt gggttgaag tttgaggtgt	240
tcaacggaag tgctccttct ttgttctcta ctacatattc atggacaaag atggcaaaca	300
ttatattctt ggatcagcca gtaggatctg gcttctccta ctcaaaaact ccaattgata	360
aaactggta cataagtgaa gtaaagagga cccatgagtt tcttcaaaag tggctaagca	420
ggcatccaca atatttctcc aaccctttat atgtagttgg agattcttat tccggtatga	480
ttgtcccggc cctcgttcaa gaaatctcac aaggaaatta tatatgttgc gaacctccta	540
taaatctaca gggttatatg cttggaaacc ctgtaacata tatggacttt gaacaaaact	600
tccgcattcc atatgcttat ggtatggat taatctccga cgaaatctat gagccatgaa	660
agagaatctg caacggaaat tattacaatg tggatccatc taacacacaa tggttggaaac	720
ttactgaaga atatcataag tgcactgccaa aataaaatat ccatcacata ttaacaccag	780
attgcgatgt aaccaatgta acatctcctg attgttatta ttatccatat catctcattt	840
aatgttgggc taacgacgag agcggtcgcg aagctttca tattgaaaag ggtagtaaag	900
gaaaatgggc gcgatgtaat cgactattc catacatca cgacattgta agcagcatac	960
catatcacat gaataacacgc atcagtggat accgatctct tatttacagt ggtgatcagc	1020
acatcgccgt ccctttctt gcaactcaag cctggataag atctctcaat tactccccca	1080
ttcataactg gaggccatgg atgataaaca atcaaatcgc tggatacacg agagcttatt	1140
ccaataagat gacatttgct actatcaaag gaggtggaca cacggcagag tatagaccaa	1200
acgagacctt tatcatgttc caaagggtgga tcagtgccaa acccctgtaa gtcgac	1256
<210> 7	
<211> 415	

<212> PRT

<213> Arabidopsis thaliana

<400> 7

Met Thr Ser Ile Val Lys Phe Leu Pro Gly Phe Glu Gly Pro Leu Pro
1 5 10 15

Phe Glu Leu Glu Thr Gly Tyr Ile Gly Ile Gly Glu Asp Glu Asn Val
20 25 30

Gln Phe Phe Tyr Tyr Phe Ile Lys Ser Glu Asn Asn Pro Lys Glu Asp
35 40 45

Pro Leu Leu Ile Trp Leu Asn Gly Gly Pro Gly Cys Ser Cys Leu Gly
50 55 60

Gly Ile Ile Phe Glu Asn Gly Pro Val Gly Leu Lys Phe Glu Val Phe
65 70 75 80

Asn Gly Ser Ala Pro Ser Leu Phe Ser Thr Thr Tyr Ser Trp Thr Lys
85 90 95

Met Ala Asn Ile Ile Phe Leu Asp Gln Pro Val Gly Ser Gly Phe Ser
100 105 110

Tyr Ser Lys Thr Pro Ile Asp Lys Thr Gly Asp Ile Ser Glu Val Lys
115 120 125

Arg Thr His Glu Phe Leu Gln Lys Trp Leu Ser Arg His Pro Gln Tyr
130 135 140

Phe Ser Asn Pro Leu Tyr Val Val Gly Asp Ser Tyr Ser Gly Met Ile
145 150 155 160

Val Pro Ala Leu Val Gln Glu Ile Ser Gln Gly Asn Tyr Ile Cys Cys
165 170 175

Glu Pro Pro Ile Asn Leu Gln Gly Tyr Met Leu Gly Asn Pro Val Thr
180 185 190

Tyr Met Asp Phe Glu Gln Asn Phe Arg Ile Pro Tyr Ala Tyr Gly Met
195 200 205

Gly Leu Ile Ser Asp Glu Ile Tyr Glu Pro Met Lys Arg Ile Cys Asn
210 215 220

Gly Asn Tyr Tyr Asn Val Asp Pro Ser Asn Thr Gln Cys Leu Lys Leu
225 230 235 240

Thr Glu Glu Tyr His Lys Cys Thr Ala Lys Ile Asn Ile His His Ile
245 250 255

Leu Thr Pro Asp Cys Asp Val Thr Asn Val Thr Ser Pro Asp Cys Tyr
260 265 270

Tyr Tyr Pro Tyr His Leu Ile Glu Cys Trp Ala Asn Asp Glu Ser Val
275 280 285

Arg Glu Ala Leu His Ile Glu Lys Gly Ser Lys Gly Lys Trp Ala Arg
290 295 300

Cys Asn Arg Thr Ile Pro Tyr Asn His Asp Ile Val Ser Ser Ile Pro
305 310 315 320

Tyr His Met Asn Asn Ser Ile Ser Gly Tyr Arg Ser Leu Ile Tyr Ser
325 330 335

Gly Asp His Asp Ile Ala Val Pro Phe Leu Ala Thr Gln Ala Trp Ile
340 345 350

Arg Ser Leu Asn Tyr Ser Pro Ile His Asn Trp Arg Pro Trp Met Ile
355 360 365

Asn Asn Gln Ile Ala Gly Tyr Thr Arg Ala Tyr Ser Asn Lys Met Thr
370 375 380

Phe Ala Thr Ile Lys Gly Gly Gly His Thr Ala Glu Tyr Arg Pro Asn
385 390 395 400

Glu Thr Phe Ile Met Phe Gln Arg Trp Ile Ser Gly Gln Pro Leu
405 410 415

<210> 8

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<223> primer

<400> 8

gagaatatca tgagtttgaa aataaag

27

<210> 9

<211> 1317

<212> DNA

<213> Arabidopsis thaliana

<400> 9

gagaatatca tgagtttgaa aataaagttt ctgcttctgc ttgtcttgta tcatcatgtt

60

gattctgcct ctatcgtaa gtttcttcct ggaaaaaagggac gcccctttcc tttcgaacctt

120

gaaaaccgggt acattggtat tggtgaggac gagaatgtgc aatttttcta ctatccatc

180

aaatctgaaa acaatccaaa agaagatcct cttcttataat ggttaaatgg aggacctgga

240

tgttcttgta ttgggttat tatttttag aacggaccgg tgggttgaa gtttgaggtg

300

ttcaacggaa	gtgctccttc	tttgttctct	actacatatt	catggacaaa	gatggcaa	ac	360
attatattct	tggatcagcc	agtaggatct	ggcttctcct	actcaaaaac	tccaattgat		420
aaaactggtg	acataagtga	agtaaagagg	accatgagt	ttcttcaaaa	gtggcta	agc	480
aggcatccac	aatatttctc	caacccttta	tatgttagttg	gagattctta	ttccggtat	g	540
attgtcccg	ccctcggtca	agaaatctca	caaggaaatt	atatatgtt	cgaacctc	cct	600
ataaaatctac	agggttat	at	gcttgaaac	cctgtAACAT	atatggactt	tgaacaaaac	660
ttccgcattc	catatgctta	tggtatgg	ttaatctcg	acgaaatcta	tgagccat	g	720
aagagaatct	gcaacggaaa	ttattacaat	gtggatccat	ctaACACACA	atgtttgaaa		780
cttactgaag	aatatcataa	gtgcactgcc	aaaataaata	tccatcacat	attaacacca		840
gattgcgt	taaccaatgt	aacatctcct	gattgttatt	attatccata	tcatctcatt		900
gaatgttgg	ctaacgacga	gagcgttgc	gaagcttcc	atattgaaaa	gggtagtaaa		960
ggaaaatgg	cgcgatgtaa	tcggactatt	ccatacaatc	acgacattgt	aagcagcata		1020
ccatatcaca	tgaataaacag	catcagtgg	taccgatctc	ttatttacag	tggtgatcac		1080
gacatcg	cccccccc	tgcaactcaa	gcctggataa	gatctctcaa	ttactcccc		1140
attcataact	ggaggccat	gatgataaac	aatcaa	atcg	ctggatacac	gagagctt	1200
tccaataaga	tgacatttgc	tactatcaa	ggaggtggac	acacggcaga	gtatagacca		1260
aacgagac	ttatcatgtt	ccaaagg	ttgg	atcagtggcc	aacccctgta	agtcgac	1317

<210> 10

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<223> primer

<400> 10

ctactcattt catatgtcac accccgcgtt aa

32

<210> 11

<211> 34

```

<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<223> primer

<400> 11
catcttacta gatctttagt acaacgggtga cgcc          34

<210> 12
<211> 495
<212> DNA
<213> Escherichia coli

<400> 12
atgtcacacc ccgcgttaac gcaactgcgt gcgcgtcgct attgtaaaga gatccctgcc   60
ctggatccgc aactgctcga ctggctgttg ctggaggatt ccatgacaaa acgttttcaa 120
cagcagggaa aaacggtaag cgtgacgatg atccgcgaag ggtttgcga gcagaatgaa 180
atccccgaag aactgccgct gctgccgaaa gagtctcggt actggttacg tgaaattttg 240
ttatgtgccg atggtaacc gtggcttgcc ggtcgtaaccg tcgttcctgt gtcaacgtta 300
agcggggccgg agctggcggt acaaaaattg ggtaaaacgc cgtaggacg ctatctgttc 360
acatcatcga cattaacccg ggactttatt gagataggcc gtgatgccgg gctgtggggg 420
cgacgttccc gcctgcgatt aagcggtaaa ccgctgttgc taacagaact gttttaccg 480
gcgtcaccgt tgtac          495

<210> 13
<211> 165
<212> PRT
<213> Escherichia coli

<400> 13
Met Ser His Pro Ala Leu Thr Gln Leu Arg Ala Leu Arg Tyr Cys Lys
1                      5                      10                     15

```

Glu	Ile	Pro	Ala	Leu	Asp	Pro	Gln	Leu	Leu	Asp	Trp	Leu	Leu	Leu	Glu
				20				25					30		
Asp	Ser	Met	Thr	Lys	Arg	Phe	Glu	Gln	Gln	Gly	Lys	Thr	Val	Ser	Val
		35					40				45				
Thr	Met	Ile	Arg	Glu	Gly	Phe	Val	Glu	Gln	Asn	Glu	Ile	Pro	Glu	Glu
		50					55				60				
Leu	Pro	Leu	Leu	Pro	Lys	Glu	Ser	Arg	Tyr	Trp	Leu	Arg	Glu	Ile	Leu
		65			70				75				80		
Leu	Cys	Ala	Asp	Gly	Glu	Pro	Trp	Leu	Ala	Gly	Arg	Thr	Val	Val	Pro
		85						90				95			
Val	Ser	Thr	Leu	Ser	Gly	Pro	Glu	Leu	Ala	Leu	Gln	Lys	Leu	Gly	Lys
		100					105					110			
Thr	Pro	Leu	Gly	Arg	Tyr	Leu	Phe	Thr	Ser	Ser	Thr	Leu	Thr	Arg	Asp
		115					120					125			
Phe	Ile	Glu	Ile	Gly	Arg	Asp	Ala	Gly	Leu	Trp	Gly	Arg	Arg	Ser	Arg
		130				135					140				
Leu	Arg	Leu	Ser	Gly	Lys	Pro	Leu	Leu	Leu	Thr	Glu	Leu	Phe	Leu	Pro
		145			150					155			160		
Ala	Ser	Pro	Leu	Tyr											
		165													
<210>	14														
<211>	39														
<212>	DNA														
<213>	Artificial Sequence														
<220>															
<221>	misc_feature														
<223>	primer														
<400>	14														
	ctactcaattt	agatctccat	ggcttcctct	gtcatttct											39
<210>	15														
<211>	32														
<212>	DNA														
<213>	Artificial Sequence														

```

<220>

<221> misc_feature

<223> primer

<400> 15
catcttactc atatgccaca cctgcatgca gc 32

<210> 16

<211> 684

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<223> open reading frame of the chloroplast-targeted CPL fusion protein

<400> 16
atggcttcct ctgtcatttc ttcagcagct gttgccacac gcagcaatgt tacacaagct 60
agcatggttg caccttcac tggtctcaaa tcttcagcca cttccctgt tacaaagaag 120
caaaaaccttg acatcacttc cattgcttagc aatggtgaa gagtttagctg catgcaggtg 180
tggcatatgt cacaccccgca gttAACGCAA ctgcgtgcgc tgcgctattg taaagagatc 240
cctgccctgg atccgcaact gctcgactgg ctgttgctgg aggattccat gacaaaacgt 300
tttgaacagc agggaaaaac ggtaagcgtg acgatgatcc gcgaagggtt tgtcgagcag 360
aatgaaatcc ccgaagaact gccgctgctg ccgaaagagt ctcgttactg gttacgtgaa 420
attttgttat gtgccatgg tgaaccgtgg cttgccggtc gtaccgtcgt tcctgtgtca 480
acgttaagcg ggccggagct ggcgttacaa aaattggta aaacgcccgtt aggacgctat 540
ctgttcacat catgacatt aacccgggac tttattgaga taggccgtga tgccgggctg 600
tgggggcgac gttcccgccct gcgattaagc ggtAAACCGC tgttgctaac agaactgttt 660
ttaccggcgt caccgttata ctaa 684

<210> 17

<211> 227

<212> PRT

<213> Artificial Sequence

```

<220>
<221> VARIANT
<222> (1) .. (227)
<223> open reading frame of the chloroplast-targeted CPL fusion protein

<400> 17

Met Ala Ser Ser Val Ile Ser Ser Ala Ala Val Ala Thr Arg Ser Asn
1 5 10 15

Val Thr Gln Ala Ser Met Val Ala Pro Phe Thr Gly Leu Lys Ser Ser
20 25 30

Ala Thr Phe Pro Val Thr Lys Lys Gln Asn Leu Asp Ile Thr Ser Ile
35 40 45

Ala Ser Asn Gly Gly Arg Val Ser Cys Met Gln Val Trp His Met Ser
50 55 60

His Pro Ala Leu Thr Gln Leu Arg Ala Leu Arg Tyr Cys Lys Glu Ile
65 70 75 80

Pro Ala Leu Asp Pro Gln Leu Leu Asp Trp Leu Leu Leu Glu Asp Ser
85 90 95

Met Thr Lys Arg Phe Glu Gln Gln Gly Lys Thr Val Ser Val Thr Met
100 105 110

Ile Arg Glu Gly Phe Val Glu Gln Asn Glu Ile Pro Glu Glu Leu Pro
115 120 125

Leu Leu Pro Lys Glu Ser Arg Tyr Trp Leu Arg Glu Ile Leu Leu Cys
130 135 140

Ala Asp Gly Glu Pro Trp Leu Ala Gly Arg Thr Val Val Pro Val Ser
145 150 155 160

Thr Leu Ser Gly Pro Glu Leu Ala Leu Gln Lys Leu Gly Lys Thr Pro
165 170 175

Leu Gly Arg Tyr Leu Phe Thr Ser Ser Thr Leu Thr Arg Asp Phe Ile
180 185 190

Glu Ile Gly Arg Asp Ala Gly Leu Trp Gly Arg Arg Ser Arg Leu Arg
195 200 205

Leu Ser Gly Lys Pro Leu Leu Leu Thr Glu Leu Phe Leu Pro Ala Ser
210 215 220

Pro Leu Tyr
225

<210> 18

<211> 34

```
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<223> primer
)

<400> 18
ctactcattt gaagactgca tgcagggttg gcat          34

<210> 19
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<223> primer

<400> 19
catcttactg tcgacttttag tacaacggtg acgc          34
```